

# 地図上の投稿情報の信頼度

## Credibility of the posted information on a map

山本 浩司\*<sup>1</sup>    片上 大輔\*<sup>1</sup>    新田 克己\*<sup>1</sup>    相場 亮\*<sup>2</sup>    桑田 仁\*<sup>2</sup>  
 Koji Yamamoto    Daisuke Katagami    Katsumi Nitta    Akira Aiba    Hitoshi Kuwata

\*<sup>1</sup>東京工業大学 大学院総合理工学研究科  
 知能システム科学専攻  
 CISS, IGSSE, Tokyo Institute of Technology

\*<sup>2</sup>芝浦工業大学システム工学部  
 Faculty of Systems Engineering, Shibaura Institute of Technology

In this report, we propose a method for measuring the credibility of the posted information from users. we are developing a system which recommends the route and destination using other user's posted information. However, posted information can include subjective information from various perspective, so we can not trust all posted information as they are. We propose and integrate two types of credibility of posted information. First, credibility of voting by other users is determined based on the posting user's own credibility and other users' credibility who approved or disapproved her/his posting information. The other is credibility of geographic posting tendency of the users. This credibility is determined based on dispersion of posted information on the map.

### 1. はじめに

近年、地図上で情報を表示したり、その情報を用いてユーザをナビゲーションするようなアプリケーションが開発されている。こういったアプリケーションでは、掲載されている情報の鮮度を保つためにいかに情報の更新を行うかが問題となるため、我々はユーザから募った投稿情報を利用して、ユーザの現在地や属性を考慮したリアルタイムのナビゲーションを行うためのシステムを搭載した知的車椅子を開発してきた。この車椅子に搭載されるシステムは、ユーザからの投稿情報を用いて情報の投稿を各ユーザから受け付けると同時に、それらの情報を利用してユーザへの情報や経路の推薦を行うものである。

この情報は一般の投稿をもとにしているため、信頼性が問題となる。従来から、投稿の信頼性評価はさまざまな研究があるが、従来手法は、地域的な偏りを考慮していない。その地域に詳しい人の投稿は、信頼できるはずであり、それを考慮した信頼度の判定方法を考える必要がある。そこで、ここでは、投稿情報の地理的偏りを考慮した信頼度の判定方法の提案を目的とする。

### 2. システムの概要

車椅子に搭載されている本システムには、(1) ユーザからの投稿情報を受け付け、地図に表示する機能、(2) 投稿されたデータを他の一般ユーザが閲覧、評価する機能、(3) 投稿された情報を用いた経路や行き先の推薦を行う機能、という3つの機能がある(図2)。

ユーザから投稿情報を集めることで、情報の鮮度の確保が期待できる。投稿情報は地図上にアイコンで表され、ユーザはそれをクリックすることで、投稿されているその地点に関する画像やコメントや障害者用の駐車場やエレベータなどの有無といった情報を見ることができる。一度投稿された情報には、投稿の



図 1: 知的車椅子

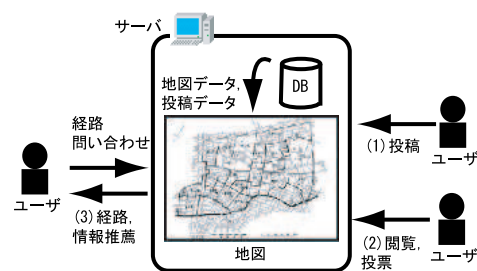


図 2: システムのイメージ

内容に同意するか、参考になったかどうかについて別のユーザが評価を追加できる。

### 3. 情報の信頼度

#### 3.1 信頼度の特性

本システムではユーザからの投稿情報を募るが、前述したように、投稿される情報すべてを直ちに信頼するわけにはいかない。ユーザからの投稿情報を用いて情報を推薦する際には、その情報が信頼するに値するものかどうかを判断する必要がある。

Web 上か実世界かにかかわらず、情報を得たときにその情

Contact: \*<sup>1</sup>〒 226-8502 横浜市緑区長津田町 4259

<http://www.ntt.dis.titech.ac.jp/>

{yamamoto,katagami,nitta }

@ntt.dis.titech.ac.jp,

\*<sup>2</sup>〒 337-8570 埼玉県さいたま市見沼区大字深作 307

{aiba,kuwata}@se.shibaura-it.ac.jp

報が確かかどうかを吟味するのに、その情報源が問題になることがある。一般にその理由は、情報の出所によって、情報の確からしさが異なってくる、と考えているからであろう。たとえば、権威のある人間の意見であれば受け入れたり信じたりしやすく、権威のない人間の意見は受け入れないし、信じにくいであろう。

本研究で想定する信頼度の特性は以下ようになる。

- 特性 1 信頼度が高い投稿を行ってきた投稿者が投稿した情報の信頼度は高い。
- 特性 2 信頼度が高い投稿を行ってきたユーザから肯定的な投票を受けた情報の信頼度は高くなる。
- 特性 3 投票された投稿の近隣にその投票者が多くの投稿をしていた場合は、その信頼度は投票者の影響を強く受ける。
- 特性 4 あるユーザの投稿が特定の地域に集中していれば、その地域はユーザが日常的に訪れる地域であり、詳しいとみなし、その周辺の情報の信頼度は高くなる (図 3)。

これらの特性に共通する性質として、信頼度の高い投稿をしてきた投稿者の行動は他の投稿者の投稿の信頼度に大きな影響を及ぼすという性質があり、言い換えれば我々は信頼度の高い投稿を行っているユーザが権威を持つという立場を取る。これらの特性を考慮し、以下で信頼度について説明する。まず、本研究における投稿情報の信頼度の算出手法の概略を以下に示す。

- 新規の投稿  $p$  が行われた場合、その投稿情報に投稿者の過去の投稿の信頼度に基づいて計算される信頼度  $R_i(\text{genre}, x_p, y_p)$  を割り当てる (1 式)。  
 $R_i(\text{genre}, x_p, y_p)$  はユーザ  $i$  の、投稿地点の座標  $(x_p, y_p)$  における、ジャンル  $\text{genre}$  に関する信頼度を表す。
- 投稿情報に投票が行われた場合、その投稿情報の信頼度  $R_i(\text{genre}, x_p, y_p)$  を、投票者の過去の投稿の信頼度と、投票を受ける情報との距離、そして投票内容 (「参考になった」「まあまあ」「いいえ」) に基づいて更新する。(3 式)。
- 投稿か投票 (1 式または 3 式) が行われたとき、地理的な投稿傾向に基づいてユーザ  $i$  の投稿情報  $p$  の信頼度に関する補正值  $geo_{ip}$  (7 式) を求め、これを用いて最終的な投稿情報の信頼度  $I_{ip}$  を求める (8 式)。

以下でこれらの処理の詳細を述べる。

### 3.2 初期信頼度の割り当て

ユーザが新規ユーザの場合など、はじめて投稿を行うときには、デフォルト値を割り当て、投稿の初期信頼度  $R_i(\text{genre}, x, y)$  を、 $R_{\text{default}}$  とする。

また、すでにユーザがそのジャンルにおける投稿を行った後、新規に投稿  $p$  が行われるときのユーザのジャンルへの信頼度  $R_i(\text{genre}, x_p, y_p)$  は、すでに投稿しているユーザの投稿情報との距離、位置関係で決定され、以下の式で求められる。

$$R_i(\text{genre}, x_p, y_p) = \frac{\sum_{q \in i \text{ の } p \text{ 以外の投稿}} w_{pq} * R_i(\text{genre}, x_q, y_q)}{\sum_{q \in i \text{ の } p \text{ 以外の投稿}} w_{pq}} \quad (1)$$

$$w_{pq} = \frac{1}{\text{dist}\{(x_p, y_p), (x_q, y_q)\}} \quad (2)$$

関数  $\text{dist}$  は 2 つの投稿情報間の距離を表す。1 式では、そのユーザの既存の投稿全体との相対的な位置関係から新しい投稿情報の初期信頼度を割り当てる。割り当てられる初期信頼度は、既存の投稿の信頼度を距離で重みづけした平均である。 $w_{pq}$  は投稿間の距離の逆数であり、この総和を投稿間の距離を相対的に扱うための正規化指数とする。これにより 1 式の分子の部分で計算される、距離で割り引いた周辺の投稿情報の信頼度の和を正規化する。これにより、新しい投稿の信頼度の初期値は、遠くにある投稿よりは、近くにある投稿の信頼度に相対的に強く影響を受けた値となる。

この式によって、新規の情報に初期値を与える処理を行い、その位置の近くにそれまでに投稿した信頼度が高い投稿が存在すれば高い初期値になる。これで特性 1 を表現する。新しく投稿された投稿情報の信頼度は、それ以前に同一ユーザが投稿した情報の信頼度によって影響を受けることになるが、我々はこのように、新しい情報が投稿されたときには、古い情報から影響を受け、その信頼度が時系列的な変化を織り込みながら変動していくという立場に立つ。ゆえに、他者からの評価がまだない投稿情報であっても、それまでにそのユーザの他の情報が高い評価を受けていれば高い信頼度を獲得することになる。

### 3.3 投票による信頼度の更新

投稿情報を見たユーザは、それに対して投票を行うことができる。このとき、対象とする情報に賛同するかしないかの意思表示を行う。賛同する投票が追加された場合、情報の信頼度を増加させる役割を果たす。また、反対の投票が行われた場合、対象となった情報の信頼度を減少させる働きをする。

ユーザ  $j$  により投票が行われたときは、以下の 3 式により、投票された投稿の信頼度を更新する。

$$R_i(\text{genre}, x_p, y_p) \leftarrow R_i(\text{genre}, x_p, y_p) + f'(f^{-1}(R_i(\text{genre}, x_p, y_p))) * V_j \sum_{q \in \text{投票者 } j \text{ の全投稿}} \left( R_j(\text{genre}, x_q, y_q) * \exp\left(-\frac{(\text{dist}\{(x_p, y_p), (x_q, y_q)\})^2}{2\sigma_1^2}\right) \right) \quad (3)$$

$V_j$  は投票を示す値で、「参考になった」= 1, 「いいえ」= -1, 「まあまあ」= 0 である。 $\sigma_1$  は投票時の距離による影響の度合いを調整するパラメータである。この式では投票者の投稿と投票を受ける投稿の距離で減少する関数と、投票者のこれまでの投稿の積の和 ( $V_j$  より後ろの項) で、投票者の投票の影響の強さを表現している。つまり、投票を受ける投稿の付近に投票者の信頼度の高い投稿があるほど、この部分の値は大きくなる。投票が反対であれば  $V_j$  が負となるため、更新後の値は減少することになる。これにより特性 2 と特性 3 を表現する。

$f(x)$  はシグモイド関数であり、

$$f(x) = \frac{1}{1 + e^{-\alpha x}} \quad (4)$$

である。この関数の導関数は

$$f'(x) = \alpha(1 - f(x))f(x) \quad (5)$$

となり、これを更新時の増分の勾配として調整するように用いている。また、 $f(x)$  の逆関数を用いて、信頼度が  $f(x)$  の値を取るときの導関数によって勾配を求める。ここでは、シグモイ

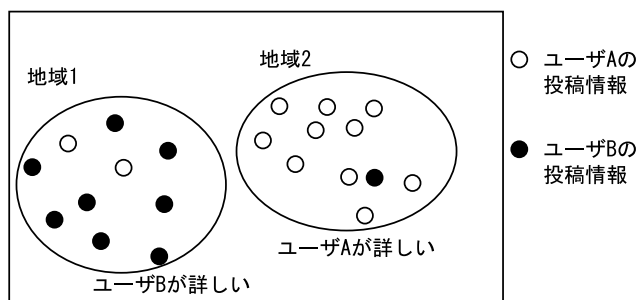


図 3: 地理的投稿傾向

ド関数の出力が信頼度の値をとるときの傾きを用いるのが目的であり、入力の  $x$  は、その出力をするときの入力値である。

$$f^{-1}(x) = -\frac{1}{\alpha} \log \frac{1-x}{x} \quad (6)$$

$\alpha$  はシグモイド関数のゲインであり、この値が大きいと勾配は急になる。

### 3.4 地理的投稿傾向による信頼度の補正

1式や3式によって、ユーザ  $i$  から投稿  $p$  がなされたとき、その信頼度  $R_i(\text{genre}, x_p, y_p)$  の初期値がどのように定義され、それが投票によってどのように更新されるかを説明した。

しかし、投稿の信頼度はそれだけでは十分ではない。その投稿の後に同一ユーザ  $i$  によってなされた投稿や、その投稿に対しての別のユーザからの投票によって、投稿  $p$  の信頼度が影響を受けるからである。そこで、投稿者  $i$  の他の投稿によって受ける影響による信頼度の補正值  $geo_{ip}$  を以下のように定義する。

$$geo_{ip} = f'(f^{-1}(R_i(\text{genre}, x_p, y_p))) * \sum_{q \in \text{ユーザ } i \text{ の } p \text{ 以外の投稿}} \left( R_i(\text{genre}, x_q, y_q) * \exp\left(-\frac{(\text{dist}\{(x_p, y_p), (x_q, y_q)\})^2}{2\sigma_2^2}\right) \right) \quad (7)$$

$\sigma_2$  は投稿情報が同一ユーザの他の投稿情報に距離に応じて与える効果を調整するパラメータである。より近傍に信頼度の高い情報が数多くある場合に、重ね合わせも大きくなるため、この効果は高まる。つまり付近にいい投稿がある場合、同じ投稿者の近くの情報もより信頼できそうだということを表現する。これは特性 4 に相当する。上述したように、投稿が投票 (1式または3式) が行われたときに、それに伴ってこの7式で信頼度の補正を行う。この値は1式のような初期値を決める時点で一度だけ組み込むのではなく、同一ユーザの他の投稿の新規投稿時や、投票を受ける時に、その時点での投稿の数や信頼度に応じて更新される。投稿が投票が行われたタイミングでこの7式で別個に計算する理由は、もしこのような効果を1式のような初期値の決定の時点において組み入れてしまうと、投稿が少ない時期にはそのような効果は少なく、投稿が増えくると大きくなるため、後に投稿した情報と先に投稿した情報においてその効果が不均一になってしまうためである。たとえば、初期値にこの効果を組み入れた場合、あるユーザが2番目の情報を投稿した時点より、5番目に投稿した時点では重ね

合わせの効果が大きくなってしまいが、2番目の情報の効果は更新されないままになってしまう。その時点ではそのユーザの投稿は5個存在しており、その時点における効果は「そのユーザは現時点で5個の投稿を行っている」という事実のもとで、どの順番で投稿されたかに関係なく計算されるべきであり、そのような状態が起こるのは望ましくない。それを防ぐため、投稿あるいは投票が行われたタイミングで、投稿者あるいは投票を受けた投稿の投稿者のすべての投稿について、7式で再計算し、すべての情報を均一に扱うために、7式で分離して計算を行っている。

最終的に、ユーザ  $i$  の投稿した投稿  $p$  の信頼度  $I_{ip}$  を以下のように定義する。

$$I_{ip} = R_i(\text{genre}, x_p, y_p) + R_i(\text{genre}, x_p, y_p) * geo_{ip} \quad (8)$$

ここでは、その情報自身の信頼度を重視するために、周囲からの重ね合わせとその地点での信頼度の積を取ってから加えている。

上述したように、投稿が行われたときは、1式 7式 8式の順に更新が行われる。また、投票が行われたときは、3式 7式 8式の順に更新が行われる。

まとめると、手続きは以下ようになる。

- |  |
|--|
| <p>投稿が行われたとき</p> <ol style="list-style-type: none"> <li>1式で新規投稿への初期信頼度を決定</li> <li>投稿者のすべての投稿に関して7式で <math>geo_{ip}</math> を計算</li> <li>投稿者のすべての投稿に関して8式で最終的な信頼度 <math>I_{ip}</math> を計算</li> </ol> <p>投票が行われたとき</p> <ol style="list-style-type: none"> <li>3式で投票を受けた投稿の信頼度を更新</li> <li>投票を受けた投稿を行った投稿者のすべての投稿に関して7式で <math>geo_{ip}</math> を計算</li> <li>投票を受けた投稿を行った投稿者のすべての投稿に関して8式で最終的な信頼度 <math>I_{ip}</math> を計算</li> </ol> |
|--|

## 4. 実験と考察

### 4.1 実験の目的

本システムを用いて、信頼度が我々の目標とした性質を持っているかを確認するために、評価実験を行った。まず、実験の準備として投稿情報を収集した。

### 4.2 投稿情報の収集

大学院生と職員の計15名に、本システムへの情報の投稿を依頼した。地域としては、青葉台 (横浜市青葉区)、渋谷、大岡山、町田の4箇所とし、これらの地域に関して、知っている情報を投稿してもらった。投稿するためには、Webブラウザを用いて本システムにアクセスしてもらい、全員にユーザ登録を行ってもらった。それから、情報の投稿を行ってもらい、さらに、他のユーザの投稿した情報を閲覧して、評価を行ってもらった。また、本システムは、利用者同士が知り合いではないような一般的なWeb上での利用を想定している。ここでは、被験者同士が知り合いであるためのバイアスがかからないようにするため、被験者に使ってもらうユーザIDは、正

体が誰なのか推測できないような文字列を割り当てることに注意した。

パラメータは,  $R_{default} = 0.2$ ,  $\alpha = 0.2$  とした。また, 近傍関数についての  $\sigma_1, \sigma_2$  については, 4 箇所の地域のスケールを考慮し,  $\sigma_1 = 30 \text{ km}$ ,  $\sigma_2 = 15 \text{ km}$  とした。  $\sigma_2$  は投稿情報が同じ投稿者の周囲の投稿情報に与える影響を決めるが, 例えば, 渋谷に投稿が多いからといって, 町田に詳しいとは必ずしもいえないため, 投票時の投票者の影響を示す  $\sigma_1$  より小さな値としている。

134 件の投稿情報があり, 投票による評価は 412 件であった。評価は, "はい" が 179 件, "まあまあ" が 193 件, "いいえ" が 40 件であった。情報を投稿せずに, 評価だけを行ったユーザもあり, 投稿を行ったユーザは 11 人であった。

### 4.3 実験の設定

以下の 3 条件を用い, 大学院生 14 人に, すべての投稿情報について, 信頼できるかどうかを 7 点法で判定してもらった。被験者にはモデルにおける計算の詳細についての情報は与えていない。また, この判定は新たな被験者に依頼し, 判定に用いられるデータに, 上述した投稿情報の収集時の 15 人との重なりはない。

条件 1 情報のみが表示され, 個々の投稿の投稿者 ID は見えない。誰が投稿したかを知ることはできないため, 個々の記事を別個に評価することになる。

条件 2 個々の投稿に投稿者 ID を付加する。従って, 誰がどの記事を書いたかを把握することができる。

条件 3 個々の投稿に投稿者 ID と地図上の位置を示す。誰がどの記事を書いただけでなく, どの位置に投稿しているかが把握できる。また, 誰から投票を受けているかも開示する。

すべての条件において, 情報の内容自体は同じものである。条件が後になるにつれて被験者に与えられる情報が増え, 条件 3 では本システムで想定する情報 (投稿者, 投稿の位置, 投票者名とその内容) がすべて開示される。したがって, もし条件によって被験者の判定傾向が異なり, 我々のモデルで算出される信頼度が, 人間の判定した値の傾向により近いということになれば, 人間がすべての情報をチェックせずともシステムがそれに近い信頼度を自動的に算出することを人間の代わりに行うことができることになる。また, モデルの特徴として,

- 投稿数の多いユーザの情報に地理的な投稿傾向の効果による信頼度を付加すること
- 投票において重みに差があること (権威に差があること)

という点に着目し, これらの有無による違いを調べる。投稿傾向による効果をなしにするときは, 8 式で  $geo_{ip} = 0$  とした。また, 投票の重みを一律にする場合には, 3 式で  $\sum_{\text{投票したユーザ } j \text{ の投稿 } \ni p} R_j(\text{genre}, x_p, y_p)$  の部分に 0.2 を代入する。これで投票者の投稿の数や信頼度の高さに関わらず投票の重みは等しくなる。また,  $\exp\left(-\frac{(\text{dist}\{(x_p, y_p), (x_q, y_q)\})^2}{2\sigma^2}\right)$  の代わりに 1.0 を代入した式を用いる。これにより距離の影響を無くす。

### 4.4 結果と考察

4.4.1 被験者の判定とモデルとの比較各条件における, 4 章で述べたモデルの式を変更した 4 つの場合の結果を図 7 と表 1 に示す。条件 1 より条件 2 が, 条件 2 より条件 3 のほうが相関が高い。これは, 条件 2 では投稿者が誰かを把握できるようになったために, その投稿者の投稿一覧からその投稿者がど

表 1: モデルと被験者の判定の平均との相関係数

	投稿傾向効果/投票の重みの差			
	なし/なし	なし/あり	あり/なし	あり/あり
条件 1	0.432	0.435	0.427	0.430
条件 2	0.454	0.460	0.501	0.509
条件 3	0.700	0.699	0.732	0.731

表 2: モデルと被験者の判定の平均のスピアマン順位相関係数

	投稿傾向効果/投票の重みの差			
	なし/なし	なし/あり	あり/なし	あり/あり
条件 1	0.460	0.446	0.399	0.403
条件 2	0.539	0.541	0.542	0.549
条件 3	0.761	0.750	0.777	0.780

のような人物かをイメージするようになった結果, 判定がモデルに近づいたと考えられる。被験者に行ったアンケートでは, 条件 1 のときは情報自体の内容で評価するしかいないため, 「内容が具体的かどうかで評価した」とか, 「主観でなく客観的事実が書かれているかで評価した」といった意見が見られたのに対し, 条件 2 においては「どのような人物かを想像するようになった」とか「条件 1 では信頼できなさそうな情報でも, 条件 2 で名前が出て, その人の他の情報を見て信頼するようになった」といった意見が見られ, ほぼ全員が条件ごとに評価が変わったと回答した。また条件 3 では「他人の投票を参考にするようになった」という意見が多く見られた。投稿傾向による効果の有無と, 投票の重みが一律か, または投票者の信頼度を反映させるかの場合, どちらにおいてもない場合よりある場合のほうが相関が高い傾向が見られた。表 3 に各条件における被験者の評価の平均と, モデルの算出した信頼度の値の一部を示す。モデルの出力する信頼度は被験者の判定の 7 点法とは尺度が異なるため, 表 3 の右側に順位での比較を併記する。被験者の判定とモデルの順位がどのくらい一致するかを, スピアマンの順位相関係数を用いて調べたところ (表 2, 図 8), 条件 1 では 0.403, 条件 2 では 0.549, 条件 3 では 0.780 であり, 条件 3 において被験者の判定とかなり高い相関を示した。

## 5. まとめ

本稿では, 地図上の情報推薦システムにおける投稿情報の信頼度について述べた。今後, 情報の信頼度を用いたユーザ間協調の枠組みを構築する予定である。

## 参考文献

- [Pazzani 99] Michael Pazzani: A Framework for Collaborative, Content-Based and Demographic Filtering, *Artificial Intelligence Review*, pp. 393-408, 1999.
- [Vozalis 04] M. G. Vozalis and K. G. Margaritis: Collaborative Filtering enhanced by Demographic Correlation", *AIAI Symposium on Professional Practice in AI, of the 18th World Computer Congress*, 2004.
- [Zacharia 99] G. Zacharia, A. Moukas, P. Maes: Collaborative Reputation Mechanisms in Electronic Marketplaces, *Proceedings of the 32nd Hawaii International Conference on System Sciences*, 1999.